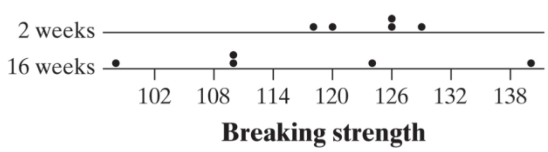
Section 10.2

Check Your Understanding, Page 644:

1. *State:* Our parameters of interest are μ_1 =the true mean price of wheat in July and μ_2 = the true mean price of wheat in September. We want to estimate the difference $\mu_1 - \mu_2$ at a 99% confidence level. *Plan:* We should use a two-sample *t* interval for $\mu_1 - \mu_2$ if the conditions are met. Random: The data come from independent random samples. 10%: n_1 = 90 is less than 10% of all wheat producers in July and n_2 = 45 is less than 10% of all wheat producers in September. Normal/Large Sample: $n_1 = 90 \geq 30$ and $n_2 = 45 \geq 30$. *Do:* The conservative df = 45 – 1 = 44. Using Table B and df = 40, the 99% confidence interval is $(2.95-3.61) \pm 2.704\sqrt{\frac{(0.22)^2}{90}+\frac{(0.19)^2}{45}} = -0.66 \pm 0.099 = (-0.759, -0.561)$. *Using technology*: (–0.756, –0.564) with df = 100.45. *Conclude:* We are 99% confident that the interval from –0.756 to –0.564 captures the true difference in mean wheat prices in July and September.

Check Your Understanding, page 649:

1. *State:* We want to perform a test at the $\alpha = 0.05$ significance level of $H_0: \mu_1 - \mu_2 = 0$ versus $H_a: \mu_1 - \mu_2 > 0$ where μ_1 is the true mean breaking strength for polyester fabric buried for 2 weeks and μ_2 is the true mean breaking strength for polyester fabric buried for 16 weeks. *Plan:* We should use a two-sample *t* test if the conditions are met. Random: The data come from two groups in a randomized experiment. Normal/Large Sample: Because the number of observations in both groups is less than 30, we must graph the data. The dotplots below show no strong skewness or outliers in either group, so it is reasonable to use a two-sample *t* procedure.



Do: From the data we find that $n_1 = 5$, $\bar{x}_1 = 123.8$, $s_1 = 4.60$, $n_2 = 5$, $\bar{x}_2 = 116.4$, and $s_2 = 16.09$. The test statistic is $t = \frac{(123.8-116.4)-0}{\sqrt{\frac{(4.60)^2}{5}+\frac{(16.09)^2}{5}}} = 0.989$. Using Table B and df = 4, the *P*-value is between 0.15 and 0.20. Using technology: *P*-value = 0.1857 with df = 4.65. *Conclude:* Because the *P*-value of 0.1857 is greater than α = 0.05, we fail to reject H_0. We do not have convincing evidence that the true mean breaking strength of polyester fabric that is buried for 2 weeks is greater than the true mean breaking strength for polyester fabric that is buried for 16 weeks.

Exercises, page 654:

10.31 (a) Let *M* = cholesterol level of a randomly selected 20- to 34-year-old male and *B* = cholesterol level of a randomly selected 14-year-old boy. Because the distributions of *M* and *B* are Normal, the distribution of $\bar{x}_M - \bar{x}_B$ is also Normal.
(b) $\mu_{\bar{x}_M - \bar{x}_B} = \mu_M - \mu_B = 188 - 170 = 18$mg/dl

(c) Because 25 is less than 10% of all 20- to 34-year old males and 36 is less than 10% of all 14-year-old boys, $\sigma_{\bar{x}_M-\bar{x}_B} = \sqrt{\frac{\sigma_M^2}{n_M}+\frac{\sigma_B^2}{n_B}} = \sqrt{\frac{(41)^2}{25}+\frac{(30)^2}{36}} = 9.60\text{ mg/dl}$

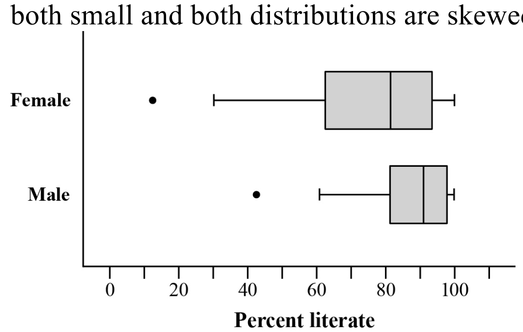
10.32 (a) Let M = the height of a randomly selected young man and W = the height of a randomly selected young woman. Because the distributions of M and W are Normal, the distribution of $\bar{x}_M - \bar{x}_W$ is also Normal.
(b) $\mu_{\bar{x}_M-\bar{x}_W} = \mu_M - \mu_W = 69.3 - 64.5 = 4.8$ inches.
(c) Because 16 is less than 10% of all young men and 9 is less than 10% of all young women,

$$\sigma_{\bar{x}_M-\bar{x}_W} = \sqrt{\frac{\sigma_M^2}{n_M}+\frac{\sigma_W^2}{n_W}} = \sqrt{\frac{(2.8)^2}{16}+\frac{(2.5)^2}{9}} = 1.09\text{ inches.}$$

10.33 The Random condition is met because these are two independent random samples. The 10% condition is met because 20 is less than 10% of all males at the school and 20 is less than 10% of all females at the school. The Normal/Large Sample condition is not met for these data. There are fewer than 30 observations in each group and the stemplot for Males shows several outliers.

10.34 The Random condition is met because these are two independent random samples. The 10% condition is met because 50 is less than 10% of students in the United Kingdom and 50 is less than 10% of students in South Africa. The Normal/Large Sample condition is met because $50 \geq 30$ and $50 \geq 30$, even though there is an outlier in the South African distribution.

10.35 The Random condition is not met because these data are not from two *independent* random samples. Knowing the literacy percent for females in a country helps us predict the literacy percent for males in that country. The 10% condition is not met because 24 is more than 10% of Islamic countries. The Normal/Large Sample condition is not met because the sample sizes are both small and both distributions are skewed to the left and have an outlier (see boxplots below).



10.36 The Random condition was not met because the words chosen from each article were the first words in the article, not a random sample of words. The 10% condition is met because 400 is less than 10% of the words in a medical journal and 100 is less than 10% of the words in an airline magazine. The Normal/Large Sample condition is met because $100 \geq 30$ and $400 \geq 30$.

10.37 (a) The distributions of percent change are both slightly skewed to the left. The centers of the two distributions seem to be quite different with people drinking red wine generally having more polyphenols in their blood. The distribution of percent change for the white wine drinkers is a little bit more variable.

(b) *State:* Our parameters of interest are μ_1 =the true mean change in polyphenol level in the blood of people like those in the study who drink red wine and μ_2 = the true mean polyphenol level in the blood of people like those in the study who drink white wine. We want to estimate the difference $\mu_1 - \mu_2$ at a 90% confidence level. *Plan:* We should use a two-sample *t* interval for $\mu_1 - \mu_2$ if the conditions are met. Random: The data come from two groups in a randomized experiment. Normal/Large Sample: The dotplots given in the problem do not show strong skewness or outliers, so it is reasonable to use two-sample *t* procedures. *Do:* From the data we find that $n_1 = 9$, $\bar{x}_1 = 5.5$, $s_1 = 2.517$, $n_2 = 9$, $\bar{x}_2 = 0.23$, and $s_2 = 3.292$. Using the conservative degrees of freedom (df = 9 – 1 = 8), the 90% confidence interval is

$$(5.5 - 0.23) \pm 1.860\sqrt{\frac{(2.517)^2}{9} + \frac{(3.292)^2}{9}} = 5.27 \pm 2.569 = (2.701, 7.839).$$ *Using technology*: (2.845, 7.689) with df = 14.97. *Conclude:* We are 90% confident that the interval from 2.845 to 7.689 captures the true difference in mean change in polyphenol level for men like these who drink red wine and men like these who drink white wine.

(c) Because all of the plausible values in the interval are positive, this interval supports the researcher's belief that red wine is more effective than white wine.

10.38 (a) The distribution of length for the red flowers is skewed to the right while the distribution of length for the yellow flowers is roughly symmetric. The center of the distribution of length is much greater for the red flowers than for the yellow flowers. The lengths of the red flowers are more variable than the lengths of the yellow flowers.

(b) *State:* Our parameters of interest are μ_1 =the true mean length of red flowers and μ_2 = the true mean length of yellow flowers. We want to estimate the difference $\mu_1 - \mu_2$ at a 95% confidence level. *Plan:* We should use a two-sample *t* interval for $\mu_1 - \mu_2$ if the conditions are met. Random: The data come from two independent random samples. 10% n_1 = 23 is less than 10% of all red flowers and n_2 = 15 is less than 10% of all yellow flowers. Normal/Large Sample: Both sample sizes were less than 30. However, the dotplots given in the problem do not indicate strong skewness or outliers, so it is reasonable to use a two-sample *t* procedure. *Do:* From the data we find that $n_1 = 23$, $\bar{x}_1 = 39.698$, $s_1 = 1.786$, $n_2 = 15$, $\bar{x}_2 = 36.18$, and $s_2 = 0.975$. Using the conservative degrees of freedom (15 – 1 = 14), the 95% confidence interval is

$$(39.698 - 36.18) \pm 2.145\sqrt{\frac{(1.786)^2}{23} + \frac{(0.975)^2}{15}} = 3.518 \pm 0.964 = (2.554, 4.482).$$ Using technology: (2.606, 4.431) with df = 35.16. *Conclude:* We are 95% confident that the interval from 2.606 to 4.431 captures the true difference in mean length of red flowers and yellow flowers.

(c) Because 0 is not in this interval, this interval does support the researchers' belief that the two varieties have different lengths.

10.39 (a) The distributions are skewed to the right because the earnings amounts cannot be negative, yet the standard deviation is almost as large as the distance between the mean and 0. The use of the two-sample *t* procedures is still justified because the sample sizes are both very large (675 ≥ 30 and 621 ≥ 30).

(b) *State:* Our parameters of interest are μ_1 = the true mean summer earnings of male students and μ_2 = the true mean summer earnings of female students. We want to estimate the difference $\mu_1 - \mu_2$ at a 90% confidence level. *Plan:* We should use a two-sample *t* interval for $\mu_1 - \mu_2$ if the conditions are met. Random: Even though the data came from a single random sample, it is reasonable to consider the two samples independent because knowing the response of a male shouldn't help us predict the response of a female. 10%: n_1 = 675 is less than 10% of male students at a large university and n_2 = 621 is less than 10% of female students at a large university. Normal/Large Sample: $n_1 = 675 \geq 30$ and $n_2 = 621 \geq 30$. *Do:* The conservative degrees of freedom is 621 – 1 = 620. Using Table B and df = 100, our 90% confidence interval is

$$(1884.52 - 1360.39) \pm 1.660\sqrt{\frac{(1368.37)^2}{675} + \frac{(1037.46)^2}{621}} = 524.13 \pm 111.45 = (412.68, 635.58).$$

Using technology: (413.62, 634.64) with df = 1249.21. *Conclude:* We are 90% confident that the interval from \$413.62 to \$634.64 captures the true difference in mean summer earnings of male students and female students at this large university.

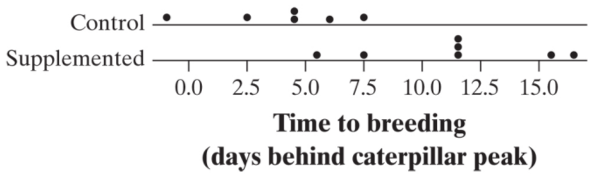
(c) It we took many random samples of 675 males and 621 females from this university and each time constructed a 90% confidence interval in this same way, about 90% of the resulting intervals would capture the true difference in mean earnings for males and females.

10.40 (a) The use of two-sample *t* procedures is still justified because both sample sizes are large ($92 \geq 30$ and $86 \geq 30$).
(b) *State:* Our parameters of interest are μ_1 = the true mean reliability rating of Anglo customers and μ_2 = the true mean reliability rating of Hispanic customers. We want to estimate the difference $\mu_1 - \mu_2$ at a 95% confidence level. *Plan:* We should use a two-sample *t* interval for $\mu_1 - \mu_2$ if the conditions are met. Random: Even though the data came from a single random sample, it is reasonable to consider the two samples independent because knowing the response of an Anglo customer shouldn't help us predict the response of a Hispanic customer. 10%: n_1 = 92 is less than 10% of Anglo customers and n_2 = 86 is less than 10% of Hispanic customers. Normal/Large Sample: $n_1 = 92 \geq 30$ and $n_2 = 86 \geq 30$. *Do:* The conservative degrees of freedom is 86 – 1 = 85. Using Table B and df = 80, the 95% confidence interval is

$$(6.37 - 5.91) \pm 1.990\sqrt{\frac{(0.60)^2}{92} + \frac{(0.93)^2}{86}} = 0.46 \pm 0.235 = (0.225, 0.695).$$ *Using technology:* (0.226, 0.694) with df = 143.69. *Conclude:* We are 95% confident that the interval from 0.226 to 0.694 captures the true difference in mean reliability rating for Anglos and Hispanics.

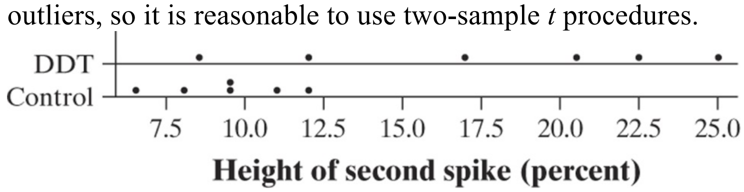
(c) If we took many random samples of 92 Anglos and 86 Hispanics and each time constructed a 95% confidence interval in this same way, about 95% of the resulting intervals would capture the true difference in mean reliability rating for Anglos and Hispanics.

10.41 (a) *State:* We want to perform a test of $H_0: \mu_1 - \mu_2 = 0$ versus $H_a: \mu_1 - \mu_2 < 0$ where μ_1 is the true mean time to breeding for the birds relying on natural food supply and μ_2 is the true mean time to breeding for birds with food-supplementation. We will use α = 0.05. *Plan:* We should use a two-sample *t* test if the conditions are met. Random: The data come from two groups in a randomized experiment. Normal/Large Sample: The comparative dotplot below displays the data for both groups. Neither distribution displays strong skewness or outliers, so it is reasonable to use two-sample *t* procedures.



Do: From the data we find that $n_1 = 6$, $\bar{x}_1 = 4.0$, $s_1 = 3.11$, $n_2 = 7$, $\bar{x}_2 = 11.3$, and $s_2 = 3.93$. The test statistic is $t = \dfrac{(4.0 - 11.3) - 0}{\sqrt{\dfrac{(3.11)^2}{6} + \dfrac{(3.93)^2}{7}}} = -3.74$. Using the conservative degrees of freedom (df = 6 – 1 = 5), the *P*-value is between 0.005 and 0.01. *Using technology*: *t* = –3.74, df = 10.95, *P*-value = 0.0016. *Conclude:* Because the *P*-value of 0.0016 is less than α = 0.05, we reject H_0. We have convincing evidence that the true mean time to breeding is less for birds relying on natural food supply than for birds with food supplements.
(b) Assuming that the true mean time to breeding is the same for birds relying on natural food supply and birds with food supplements, there is a 0.0016 probability that we would observe a difference in sample means of –7.3 or smaller by chance alone.

10.42 (a) *State:* We want to perform a test at the $\alpha = 0.05$ significance level of $H_0: \mu_1 - \mu_2 = 0$ versus $H_a: \mu_1 - \mu_2 \neq 0$ where μ_1 is the true mean relative height of the second spike in rats given DDT and μ_2 is the true mean relative height of the second spike in rats not given DDT. *Plan:* We should use a two-sample *t* test if the conditions are met. Random: These data come from two groups in a randomized experiment. Normal/Large Sample: Because both groups are small, we need to graph the data. The dotplots below show that neither group has strong skewness or outliers, so it is reasonable to use two-sample *t* procedures.



Do: From the data we find that $n_1 = 6$, $\bar{x}_1 = 17.6$, $s_1 = 6.34$, $n_2 = 6$, $\bar{x}_2 = 9.50$, and $s_2 = 1.95$. The test statistic is $t = \dfrac{(17.6 - 9.50) - 0}{\sqrt{\dfrac{(6.34)^2}{6} + \dfrac{(1.95)^2}{6}}} = 2.99$. Using the conservative degrees of freedom (df = 6 – 1 = 5), the *P*-value is between 2(0.01) and 2(0.02), so between 0.02 and 0.04. *Using technology*: *t* = 2.99, df = 5.9, *P*-value = 0.0246. *Conclude:* Because the *P*-value of 0.0246 is less than α = 0.05, we reject H_0. We have convincing evidence that the true mean relative height of the second spike is different for rats poisoned with DDT and control rats.

(b) Assuming that the true mean relative height of the second spike is the same for rats with DDT and rats without DDT, there is a 0.0246 probability of observing a difference between the means of 8.10 or greater by chance alone.

10.43 *State:* We want to perform a test at the $\alpha = 0.05$ significance level of $H_0: \mu_1 - \mu_2 = 0$ versus $H_a: \mu_1 - \mu_2 \neq 0$ where μ_1 is the true mean number of words spoken per day by female students and μ_2 is the true mean number of words spoken per day by male students. *Plan:* We should use a two-sample *t* test if the conditions are met. Random: These data come from independent random samples. 10%: $n_1 = 56$ is less than 10% of females at a large university and $n_2 = 56$ is less than 10% of males at a large university. Normal/Large Sample: $n_1 = 56 \geq 30$ and $n_2 = 56 \geq 30$. *Do:* The test statistic is $t = \dfrac{(16177 - 16569) - 0}{\sqrt{\dfrac{(7520)^2}{56} + \dfrac{(9108)^2}{56}}} = -0.248$. The conservative degrees of freedom is 56 – 1 = 55. Using Table B and df = 50, the *P*-value is greater than 2(0.25) = 0.50. *Using technology*: *t* = –0.248, df = 106.20, *P*-value = 0.8043. *Conclude:* Because the *P*-value of 0.8043 is greater than α = 0.05, we fail to reject H_0. We do not have convincing evidence that the true mean number of words spoken per day by female students is different than the true mean number of words spoken per day by male students at this university.

10.44 *State:* We want to perform a test at the $\alpha = 0.05$ significance level of $H_0: \mu_1 - \mu_2 = 0$ versus $H_a: \mu_1 - \mu_2 > 0$ where μ_1 is the true mean knee velocity for skilled rowers and μ_2 is the true mean knee velocity for novice rowers. *Plan:* We should use a two-sample *t* test if the conditions are met. Random: These data come from independent random samples. 10%: n_1 = 10 is less than 10% of skilled rowers and $n_2 = 8$ is less than 10% of novice rowers. Normal/Large Sample: Both samples had less than 30 observations. However, we are told that there were no outliers or strong skewness, so it is reasonable to use two-sample *t* procedures. *Do:* The test statistic is $t = \dfrac{(4.182 - 3.010) - 0}{\sqrt{\dfrac{(0.479)^2}{10} + \dfrac{(0.959)^2}{8}}} = 3.16$. Using Table B and the conservative degrees of freedom (df = 8 – 1 = 7), the *P*-value is between 0.005 and 0.01. *Using technology*: *t* = 3.16, df = 9.77, *P*-value = 0.0053. *Conclude:* Because the *P*-value of 0.0053 is less than α = 0.05, we reject H_0. We have convincing evidence that the mean knee velocity is greater for skilled rowers than for novice rowers.

10.45 (a) The score distribution for the activities group is slightly skewed to the left while the score distribution for the control group is slightly skewed to the right. The center of the activities group is higher than the center of the control group. The scores in the activities group are less variable than the scores in the control group. Overall, it appears that scores are typically higher for students in the activities group.

(b) *State:* We want to perform a test at the $\alpha = 0.05$ significance level of $H_0 : \mu_1 - \mu_2 = 0$ versus $H_a : \mu_1 - \mu_2 > 0$ where μ_1 is the true mean DRP score for third grade students like the ones in the experiment who do the activities and μ_2 is the true mean DRP score for third grade students like the ones in the experiment who don't do the activities. *Plan:* We should use a two-sample *t* test if the conditions are met. Random: The data come from two groups in a randomized experiment. Normal/Large Sample: Because the number of observations in both groups is less than 30, we need to look at graphs of the data. From the boxplots we see that neither distribution displays strong skewness or outliers, so it is reasonable to use two-sample *t* procedures. *Do:* The test statistic is $t = \dfrac{(51.48 - 41.52) - 0}{\sqrt{\dfrac{(11.01)^2}{21} + \dfrac{(17.15)^2}{23}}} = 2.311$. Using the conservative degrees of freedom (df = 21 – 1 = 20), the *P*-value is between 0.01 and 0.02. *Using technology*: *t* = 2.311, df = 37.86, *P*-value = 0.0132. *Conclude:* Because the *P*-value of 0.0132 is less than α = 0.05, we reject H_0. We have convincing evidence that the true mean DRP score for third grade students like the ones in the experiment who do the activities is greater than the true mean DRP score for third grade students like the ones in the experiment who don't do the activities.
(c) Because this was a randomized controlled experiment, we can conclude that the activities caused the increase in the mean DRP score.

10.46 (a) Both distributions of BMC change are slightly skewed to the right. The center of the BMC change distribution is much smaller for breast-feeding mothers than for women that are not pregnant or lactating. The BMC changes are much more variable for breast-feeding mothers than for women that are not pregnant or lactating. Overall, it appears that breast-feeding mothers do lose bone mineral.
(b) *State:* We want to perform a test at the $\alpha = 0.05$ significance level of $H_0 : \mu_1 - \mu_2 = 0$ versus $H_a : \mu_1 - \mu_2 < 0$ where μ_1 is the true mean percent change in mineral content for breastfeeding women and μ_2 is the true mean percent change in mineral content for women who were neither pregnant nor lactating. *Plan:* We should use a two-sample *t* test if the conditions are met. Random: These data came from independent random samples. 10%: $n_1 = 47$ is less than 10% of breast-feeding mothers and $n_2 = 22$ is less than 10% of women that are not pregnant or lactating. Normal/Large Sample: Because the number of observations in the control group is less than 30, we need to look at graphs of the data. Neither boxplot displays strong skewness or outliers, so it is reasonable to use a two-sample *t* procedure. *Do:* The test statistic is $t = \dfrac{(-3.587 - 0.309) - 0}{\sqrt{\dfrac{(2.506)^2}{47} + \dfrac{(1.298)^2}{22}}} = -8.499$. Using the conservative degrees of freedom (22 – 1 = 21), the *P*-value is approximately 0. Using technology: t = –8.499, df = 66.20, *P*-value ≈ 0. *Conclude:* Because the *P*-value of approximately 0 is less than α = 0.05, we reject H_0. We have convincing evidence that breastfeeding women have a greater mean percent bone mineral loss than women who are neither pregnant nor lactating.
(c) Because this was not a randomized controlled experiment, we cannot conclude that breast-feeding is the cause of the bone mineral loss. The women who are breast-feeding may be doing other things that are causing the loss in bone mineral.

10.47 *State:* We want to estimate the difference $\mu_1 - \mu_2$ at a 95% confidence level, where μ_1 is the true mean number of words spoken per day by female students and μ_2 is the true mean number of words spoken per day by male students. *Plan:* We should use a two-sample *t* interval for $\mu_1 - \mu_2$ if the conditions are met. The conditions are met (see Exercise 10.43). *Do:* The conservative degrees of freedom is 56 – 1 = 55. Using Table B and df = 50, the 95% confidence interval is $(16{,}177 - 16{,}569) \pm 2.009\sqrt{\frac{(7520)^2}{56} + \frac{(9108)^2}{56}} = -392 \pm 3171 = (-3563, 2779)$. *Using technology*: (–3521, 2737) with df = 106.2. *Conclude:* We are 95% confident that the interval from –3521 to 2737 captures the true difference in mean number of words spoken per day by female students the mean number of words spoken per day by male students. This interval allows us to determine if 0 is a plausible value for the difference in means and also provides an interval of other plausible values for the difference in mean words spoken per day.

10.48 *State:* We want to estimate the difference $\mu_1 - \mu_2$ at a 95% confidence level, where μ_1 is the true mean relative height of the second spike in rats given DDT and μ_2 is the true mean relative height of the second spike in rats not given DDT. *Plan:* We should use a two-sample *t* interval for $\mu_1 - \mu_2$ if the conditions are met. The conditions are met (see Exercise 10.42). *Do:* Using the conservative degrees of freedom (6 – 1 = 5), the 95% confidence interval is $(17.6 - 9.50) \pm 2.571\sqrt{\frac{(6.34)^2}{6} + \frac{(1.95)^2}{6}} = 8.1 \pm 6.96 =$ (1.14, 15.06). *Using technology*: (1.46, 14.74) with df = 5.94. *Conclude:* We are 95% confident that the interval from 1.46 to 14.74 captures the true difference in mean relative height of the second spike in rats given DDT and mean relative height of the second spike in rats not given DDT. This interval allows us to determine if 0 is a plausible value for the difference in means and also provides an interval of other plausible values for the difference in mean relative height of the second spike.

10.49 (a) *State:* We want to perform a test at the $\alpha = 0.05$ significance level of $H_0: \mu_1 - \mu_2 = 10$ versus $H_a: \mu_1 - \mu_2 > 10$ where μ_1 is the true mean cholesterol reduction for people like the ones in the study when using the new drug and μ_2 is the true mean cholesterol reduction for people like the ones in the study when using the current drug. *Plan:* We should use a two-sample *t* test if the conditions are met. Random: The data come from two groups in a randomized experiment. Normal/Large Sample: Both samples had less than 30 observations, but we are told that graphs of the data show no strong skewness or outliers. Thus, a two-sample *t* procedure is appropriate. *Do:* The test statistic is $t = \frac{(68.7 - 54.1) - 10}{\sqrt{\frac{(13.3)^2}{15} + \frac{(11.93)^2}{14}}} = 0.982$. Using the conservative degrees of freedom (14 – 1 = 13), the *P*-value is between 0.15 and 0.20. *Using technology*: *t* = 0.982, df = 26.96, *P*-value = 0.1675 (*Note*: To perform this test on the calculator, use $\bar{x}_1$ = 68.7 – 10 = 58.7). *Conclude:* Because the *P*-value of 0.1675 is greater than α = 0.05, we fail to reject H_0. We do not have convincing evidence that the true mean cholesterol reduction is more than 10 mg/dl greater for the new drug than for the current drug.
(b) Because we failed to reject the null hypothesis, we could have committed a Type II error. It is possible that the difference in mean cholesterol reduction is more than 10 mg/dl greater for the new drug than the current drug, but we didn't find convincing evidence that it was.

10.50 (a) *State:* We want to perform a test at the $\alpha = 0.05$ significance level of $H_0 : \mu_1 - \mu_2 = 0.5$ versus $H_a : \mu_1 - \mu_2 > 0.5$ where μ_1 is the true mean amount of water used in the current-model toilets and μ_2 is the true mean amount of water used in the new model of toilets. *Plan:* We should use a two-sample *t* test if the conditions are met. Random: The data come from independent random samples. 10%: n_1 = 30 is less than 10% of current-model toilets and n_2 =30 is less than 10% of the new model toilets. Normal/Large Sample: $n_1 = 30 \geq 30$ and $n_2 = 30 \geq 30$. *Do:* The test statistic is $t = \dfrac{(1.64 - 1.09) - 0.5}{\sqrt{\dfrac{(0.29)^2}{30} + \dfrac{(0.18)^2}{30}}} = 0.802$. Using the conservative degrees of freedom (30 – 1 = 29), the *P*-value is between 0.20 and 0.25. *Using technology*: *t* = 0.802, df = 48.46, *P*-value = 0.2131 (*Note*: To perform this test on the calculator, use $\bar{x}_1$ = 1.64 – 0.5 = 1.14). *Conclude:* Because the *P*-value of 0.2131 is greater than α = 0.05, we fail to reject H_0. We do not have convincing evidence that the new model toilet reduces the amount of water used by greater than 0.50 gallons per flush than the current-model toilet, on average.
(b) Because we failed to reject the null hypothesis, we could have committed a Type II error. That is, it is possible that the new model of toilet does reduce the amount of water used by more than 0.50 gallons, on average, but we did not find convincing evidence of this fact.

10.51 (a) The researchers randomly assigned the subjects to create two groups that were roughly equivalent at the beginning of the experiment. In other words, the random assignment should balance out the effect of other variables among the two groups.
(b) Based on the dotplot from Fathom, a difference in means of 4.15 or greater is quite rare. Only about 5 out of the 1000 differences were that big, meaning that the *P*-value is approximately 0.005. Because the *P*-value of 0.005 is less than α = 0.05, we have convincing evidence that the true mean rating for students like these that are provided with internal reasons is higher than the true mean rating for students like these that are provided with external reasons.
(c) Because we found convincing evidence that the mean is higher for students with internal reasons when it is possible that there is no difference in the means, we could have made a Type I error.

10.52 (a) If people were allowed to choose which group they wanted to be in, it is likely that all those who choose one particular treatment (sleep deprivation, for example) might be systematically different from those who choose to be in the other treatment group. Then, we wouldn't know if the difference in characteristics or difference in sleep was the cause of a difference in times.
(b) Based on the dotplot from Fathom, a difference in means of 15.92 or greater is quite rare. Only about 5 out of the 1000 differences were that big, meaning that the *P*-value is approximately 0.005. Because the *P*-value of 0.005 is less than α = 0.05, we have convincing evidence that the mean increase in score is higher for subjects like these who are allowed to sleep than for subjects like these who are sleep deprived.
(c) Because we found convincing evidence that the mean is higher for subjects with unrestricted sleep when it is possible that there is no difference in the means, we could have made a Type I error.

10.53 (a) Two-sample *t* test. The data are being produced using two distinct groups of cars in a randomized experiment.

(b) Paired *t* test. This is a matched pairs experimental design where both treatments are applied to each subject.
(c) Two-sample *t* test. Even though there are before and after measurements for each woman in the experiment, the data are being produced using two distinct groups of women.

10.54 (a) Paired *t* test. This is a matched pairs design using pairs of pigs who were littermates. One pig in each pair received one treatment and the other pig in the pair received the other treatment.
(b) Two-sample *t* test. The samples of men and women are independent. Knowing the salaries of a man in the sample does not help us predict the salary of a woman in the other sample.
(c) Two-sample *t* test. The data are being produced using two distinct groups of plots in a randomized experiment.

10.55 (a) The appropriate test is the paired *t* test because we have paired data (two scores for each student).
(b) *State*: We want to perform a test of $H_0 : \mu_d = 0$ versus $H_a : \mu_d > 0$ where μ_d is the true mean increase in SAT verbal scores of students who were coached. We will perform that test at the $\alpha = 0.05$ significance level. *Plan*: If conditions are met, we should do a paired *t* test for μ_d. Random: The data came from a random sample. 10%: n_d = 427 is less than 10% of students who are coached for their second attempt on the SAT. Normal/Large Sample: n_d = 427 ≥ 30.
Do: The test statistic is $t = \dfrac{29-0}{59/\sqrt{427}} = 10.16$, df = 426, and the *P*-value is approximately 0.
Conclude: Because the *P*-value of approximately 0 is less than α = 0.05, we reject H_0. There is convincing evidence that students who are coached increase their scores on the SAT verbal test, on average.

10.56 (a) *State:* We want to estimate the difference $\mu_1 - \mu_2$ at a 99% confidence level where μ_1 is the true mean increase in SAT verbal scores for coached students and μ_2 is the true mean increase in SAT verbal scores for students who were not coached. *Plan:* We should use a two-sample *t* interval for $\mu_1 - \mu_2$ if the conditions are satisfied. Random: The data come from independent random samples. 10%: n_1 = 427 is less than 10% of all coached students and n_2 =2733 is less than 10% of all uncoached students. Normal/Large Sample: n_1 = 427 ≥ 30 and n_2 = 2733 ≥ 30. *Do:* The conservative degrees of freedom is 427 – 1 = 426. Using Table B and df = 100, our 99% confidence interval is $(29-21) \pm 2.626\sqrt{\dfrac{(59)^2}{427} + \dfrac{(52)^2}{2733}} = 8 \pm 7.94 = (0.06, 15.94)$.
Using technology: (0.18, 15.82) with df = 534.45. *Conclude:* We are 99% confident that the interval from 0.18 to 15.82 captures the true difference in mean increase in scores on the SAT verbal exam for students who are coached and students who are not coached.
(b) Because the interval does not include 0, there is convincing evidence to conclude that students who receive coaching have a higher mean increase in scores on the SAT verbal exam than those who do not receive coaching.
(c) The average amount of points gained (0.18 to 15.82) is not very large. It does not seem like the money spent on coaching is worth it.

10.57 a

10.58 d

10.59 b

10.60 d

10.61 (a) One-sample z interval for a proportion.
(b) Paired t test for the mean difference.
(c) Two-sample z interval for the difference in proportions.
(d) Two-sample t test for a difference in means.

10.62 (a) Paired t interval for the mean difference.
(b) One-sample z test for a proportion.
(c) Two-sample z test for a difference in proportions.
(d) Two-sample t interval for a difference in means.

10.63 (a) By the 68–95–99.7 rule, about 95% of all observations fall within the interval $\mu_{\bar{x}} - 2\sigma_{\bar{x}}$ to $\mu_{\bar{x}} + 2\sigma_{\bar{x}}$. Thus, about 5% of all observations will fall outside of this interval. P(at least one mean outside interval) = 1 – P(neither mean outside interval) = 1 – $(0.95)^2$ = 1 – 0.9025 = 0.0975.
(b) By the 68–95–99.7 rule, about 95% of all observations fall within the interval $\mu_{\bar{x}} - 2\sigma_{\bar{x}}$ to $\mu_{\bar{x}} + 2\sigma_{\bar{x}}$. Thus, about 2.5% (half of 5%) of all observations will fall above $\mu_{\bar{x}} + 2\sigma_{\bar{x}}$. Let X = the number of samples that must be taken to observe one falling above $\mu_{\bar{x}} + 2\sigma_{\bar{x}}$. Then X is a geometric random variable with p = 0.025. Thus, $P(X=4) = (1-0.025)^3(0.025) = 0.0232$.
(c) By the 68–95–99.7 rule, the probability of any one observation falling within the interval $\mu_{\bar{x}} - \sigma_{\bar{x}}$ to $\mu_{\bar{x}} + \sigma_{\bar{x}}$ is about 0.68 and the probability of falling outside this interval is about 0.32. Let X = the number of sample means out of 5 that fall outside this interval. Assuming that the samples are independent, X is a binomial random variable with n = 5 and p = 0.32. We want $P(X \geq 4) = 1 - P(X \leq 3)$ = 1 – binomcdf(trials: 5, p: 0.32, x value: 3) = 1 – 0.961 = 0.039. There is a 0.039 probability that at least 4 of the 5 sample means fall outside of this interval. This is a reasonable criteria because when the process is under control, we would only get a "false alarm" about 4% of the time.

10.64 (a) *State*: We want to estimate p = the true proportion of all adults who use the internet at a 95% confidence level. *Plan*: We should use a one-sample z interval if the conditions are met. Random: the data come from a random sample. 10%: n = 2092 is less than 10% of all adults. Large Counts: the number of successes (1318) and failures (774) are both at least 10. *Do*: For these data, $\hat{p} = \frac{1318}{2092} = 0.630$. The 95% confidence interval is

$$0.630 \pm 1.96\sqrt{\frac{0.630(0.370)}{2092}} = 0.630 \pm 0.021 = (0.609, 0.651).$$

Conclude: We are 95% confident that the interval from 0.609 to 0.651 captures the true proportion of adults who use the Internet.

(b) *State:* Our parameters of interest are p_1 = true proportion of adult Internet users who expect businesses to have Web sites and p_2 = true proportion of adult non-Internet users who expect businesses to have Web sites. We want to estimate the difference $p_1 - p_2$ at a 95% confidence level. *Plan:* We should use a two-sample z interval if the conditions are met. Random: Even though the data came from a single random sample, it is reasonable to consider the two samples independent because knowing the response of an adult Internet user shouldn't help us predict the response of an adult non-Internet user. 10%: $n_1 = 1318$ is less than 10% of adult Internet users and $n_2 = 774$ is less than 10% of adult Internet nonusers. Large Counts: Both samples have at least 10 successes and failures (Internet users: 1041 successes and 277 failures. Non-Internet users: 294 successes and 480 failures). *Do:* From the data we find that $n_1 = 1318$, $\hat{p}_1 = \frac{1041}{1318} = 0.790$, $n_2 = 774$, and $\hat{p}_2 = \frac{294}{774} = 0.380$. So the 95% confidence interval is

$$(0.790 - 0.380) \pm 1.96\sqrt{\frac{0.790(0.210)}{1318} + \frac{0.380(0.620)}{774}} = 0.41 \pm 0.041 = (0.369, 0.451).$$

Conclude: We are 95% confident that the interval from 0.369 to 0.451 captures the true difference in the proportion of Internet users and non-Internet users who expect businesses to have a web site.

10.65 (a) Perhaps the people who responded are prouder of their improvements and are more willing to share. This could lead to an overestimate of the true mean improvement.
(b) This was an observational study, not an experiment. The students (or their parents) chose whether or not to be coached; students who choose coaching might have other motivating factors that help them do better the second time. For example, perhaps students who choose coaching have some personality trait that also compels them to try harder the second time. Then, we wouldn't know if it was the coaching or the personality trait that caused the increase.